ESTIMATION OF MEANS IN GRAPHICAL GAUSSIAN MODELS WITH SYMMETRIES

By Helene Gehrmann and Steffen L. Lauritzen

University of Oxford

We study the problem of estimability of means in undirected graphical Gaussian models with symmetry restrictions represented by a colored graph. Following on from previous studies, we partition the variables into sets of vertices whose corresponding means are restricted to being identical. We find a necessary and sufficient condition on the partition to ensure equality between the maximum likelihood and least-squares estimators of the mean.

1. Introduction. The elegant principles of symmetry and invariance appear in many areas of statistical research [e.g., Dawid (1988), Diaconis (1988), Eaton (1989), Viana (2008)]. Symmetry restrictions in the multivariate Gaussian distribution have a long history [Andersson (1975), Andersson, Brøns and Jensen (1983), Jensen (1988), Olkin (1972), Olkin and Press (1969), Votaw (1948), Wilks (1946)] and have recently been combined with conditional independence relations [Andersen et al. (1995), Højsgaard and Lauritzen (2008), Hylleberg, Jensen and Ørnbøl (1993), Madsen (2000)].

This article is concerned with graphical Gaussian models with symmetry constraints introduced by Højsgaard and Lauritzen (2008). The types of restrictions are: equality between specified elements of the concentration matrix (RCON), equality between specified partial correlations (RCOR) and restrictions generated by permutation symmetry (RCOP), a special instance of the former two. The models can be represented by vertex and edge colored graphs, where parameters associated with equally colored vertices or edges are restricted to being identical.

We consider maximum likelihood estimation of a nonzero mean μ subject to specific equality constraints, assuming the covariance matrix Σ satisfies the restrictions of one of the above models. This could be relevant, for
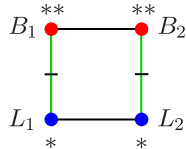
Fig. 1.   *Graphical Gaussian symmetry model supported by Frets's heads.*

example, if treatment effects are to be estimated in an experiment where the basic error structure in the units exhibits conditional independencies in a symmetric pattern.

For the Gaussian distribution, maximum likelihood estimation of μ under an unknown covariance structure is generally nontrivial, as the maximizer of the likelihood function in μ for fixed Σ may depend on Σ. The least-squares estimator μ^*, however, is defined by minimizing the sum of squares so that in case of equality of $\hat{\mu}$ and μ^*, the former is independent of Σ. Kruskal (1968) showed that for for fixed Σ, and μ in an affine space Ω, $\hat{\mu}$ and μ^* agree if and only if Ω is stable under Σ, or equivalently under $K = \Sigma^{-1}$; see also Haberman (1975) and Eaton (1983). Here we derive a necessary and sufficient condition on the graph coloring representing a model and the symmetry constraints on the mean vector μ which ensures this stability and hence equality of estimators.

We let $G = (V, E)$ denote the dependence graph of the model and let its colored version be denoted by $\mathcal{G} = (\mathcal{V}, \mathcal{E})$, where $\mathcal{V}$ denotes a partition of V into vertex color classes and $\mathcal{E}$ a partition of E into edge color classes. The symbol $\mathcal{M}$ is to denote a partition of V such that whenever two vertices are in the same set of $\mathcal{M}$, the corresponding means are restricted to being equal. The necessary and sufficient condition for equality of $\hat{\mu}$ and μ^* in the symmetry model represented by $(\mathcal{V}, \mathcal{E})$ is twofold: (i) the partition $\mathcal{M}$ must be finer than $\mathcal{V}$; and (ii) the partition must be *equitable* with respect to every edge color class in $\mathcal{E}$ as defined by Sachs (1966).

For example, the graph in Figure 1 represents a graphical Gaussian symmetry model for data concerning the head dimensions of first and second sons known as *Frets's heads* [Frets (1921), Mardia, Kent and Bibby (1979)]; here L_1, B_1 denotes the length and breadth of the head of the first son, and similarly for L_2, B_2.

This model has previously been found to be well supported by the data [Gehrmann (2011), Højsgaard and Lauritzen (2008), Whittaker (1990)] when no constraints were considered on the means. We may, for example, be interested in the hypothesis that the two lengths have the same mean, and the two breadths have the same mean, indicating that head dimensions do not generally change with the parity of the son. We shall demonstrate that this hypothesis is simple in the sense that the maximum likelihood estimator, or

MLE for short, of the mean under this hypothesis can be found by a simple average. Also, we shall demonstrate that this is not the case if we consider lengths and breadths separately.

2. Preliminaries and notation.

2.1. *Graphical Gaussian models.* Let $G = (V, E)$ be an undirected graph with vertex set V and edge set E and let $Y = (Y_\alpha)_{\alpha \in V}$ be a multivariate Gaussian random vector. Then the *graphical Gaussian model* represented by G is the set of Gaussian distributions for which Y_α is conditionally independent of Y_β given the remaining variables, denoted $Y_\alpha \perp\!\!\!\perp Y_\beta \mid Y_{V \setminus \{\alpha, \beta\}}$, whenever there is no edge between α and β in G.

For a multivariate Gaussian $\mathcal{N}_{|V|}(\mu, \Sigma)$ distribution with concentration matrix $\Sigma^{-1} = K = (k_{\alpha\beta})_{\alpha, \beta \in V}$, it holds that $Y_\alpha \perp\!\!\!\perp Y_\beta \mid Y_{V \setminus \{\alpha, \beta\}}$ if and only if $k_{\alpha\beta} = 0$. Thus letting $\mathcal{S}^+(G)$ denote the set of symmetric positive definite matrices indexed by V whose $\alpha\beta$-entry is zero whenever $\alpha\beta \notin E$, the graphical Gaussian model represented by G assumes

$$Y \sim \mathcal{N}_{|V|}(\mu, \Sigma), \qquad \mu \in \Omega = \mathbb{R}^V, \Sigma^{-1} = K \in \mathcal{S}^+(G).$$

For further details, see, for example, Lauritzen (1996), Chapter 5.

2.2. *Graph coloring.* For general graph terminology we refer to Bollobás (1998). Following Højsgaard and Lauritzen (2008) we define the following notation for graph colorings. Let $G = (V, E)$ be an undirected graph. Then a *vertex coloring* of G is a partition $\mathcal{V} = \{V_1, \ldots, V_k\}$ of V, where we refer to $V_1, \ldots, V_k$ as the *vertex color classes*. Similarly, an *edge coloring* of G is a partition $\mathcal{E} = \{E_1, \ldots, E_l\}$ of E into l *edge color classes* $E_1, \ldots, E_l$. We call a color class containing one element a *singleton* and a partition containing only singletons as elements a *singleton partition*.

Then $\mathcal{G} = (\mathcal{V}, \mathcal{E})$ denotes the *colored graph* with vertex coloring $\mathcal{V}$ and edge coloring $\mathcal{E}$; we also say that $(\mathcal{V}, \mathcal{E})$ is a *graph coloring*. We indicate the color class of a vertex by the number of asterisks we place next to it. Similarly we indicate the color class of an edge by dashes. color classes which are singletons are displayed in black and without asterisks or dashes.

2.3. *Further notation.* As we shall be considering constraints on the mean vector defined by partitions of the mean vector into groups of equal entries, we introduce the following notation. For $\mathcal{M}$ a partition of V and $\alpha, \beta \in V$, we write $\alpha \equiv \beta(\mathcal{M})$ to denote that α and β lie in the same set in $\mathcal{M}$ and let $\Omega(\mathcal{M})$ be the linear space of vectors which are constant on each set of the partition $\mathcal{M}$:

(1) $\qquad \Omega(\mathcal{M}) = \{(x_\alpha)_{\alpha \in V} \in \mathbb{R}^V : x_\alpha = x_\beta \text{ whenever } \alpha \equiv \beta(\mathcal{M})\}.$

For two partitions $\mathcal{M}_1$ and $\mathcal{M}_2$ of the same set, we shall say that $\mathcal{M}_1$ is *finer than* $\mathcal{M}_2$, denoted by $\mathcal{M}_1 \leq \mathcal{M}_2$, if every set in $\mathcal{M}_2$ can be expressed as a union of sets in $\mathcal{M}_1$. We equivalently say that $\mathcal{M}_2$ is *coarser than* $\mathcal{M}_1$.

If A is a set of edges in a graph G, for $\alpha \in V$ we shall write $\text{ne}_A(\alpha)$ to denote the set of vertices which are connected to α by an edge inside A.

We further adopt the following notation from Højsgaard and Lauritzen (2008). For a colored graph $\mathcal{G} = (\mathcal{V}, \mathcal{E})$ and $u \in \mathcal{V}$ we let T^u denote the $|V| \times |V|$ diagonal matrix with $T^u_{\alpha\alpha} = 1$ if $\alpha \in u$ and 0 otherwise. Similarly, for each edge color class $u \in \mathcal{E}$ we let T^u be the $|V| \times |V|$ symmetric matrix with $T^u_{\alpha\beta} = 1$ if $\{\alpha, \beta\} \in u$ and 0 otherwise.

3. Maximum likelihood and least-squares estimation. Letting $K = \Sigma^{-1}$ as above, the density of $Y \sim \mathcal{N}_{|V|}(\mu, \Sigma)$ is given by

$$f_{\mu,K}(y) = \frac{\det K^{1/2}}{(2\pi)^{|V|/2}} \exp\{-(y - \mu)^T K (y - \mu)/2\}$$

so that the likelihood function based on a sample $\mathcal{Y} = (Y^i)_{1 \leq i \leq n}$ where Y^i are independent, and $Y^i \sim \mathcal{N}_{|V|}(\mu, \Sigma)$ becomes

$$(2) \qquad L(\mu, K; y) \propto \det K^{n/2} \exp\left\{-\sum_{1 \leq i \leq n} (y^i - \mu)^T K (y^i - \mu)/2\right\}.$$

If μ is unrestricted, so that $\mu \in \Omega = \mathbb{R}^V$, the likelihood function L in (2) is maximized over μ for fixed K by the least squares estimator $\mu^* = \bar{y}$, and inference about K can be based on the profile likelihood function

$$(3) \qquad L(\hat{\mu}, K; y) \propto \det K^{n/2} \exp\{-\text{tr}(KW)/2\},$$

where $W = \sum_{i=1}^n (y^i - \mu^*)(y^i - \mu^*)^T$ is the matrix of sums of squares and products of the residuals. However, inference about μ when K is unknown and needs to be estimated is generally not possible, a classic example being known as the *Behrens–Fisher problem* [see Scheffé (1944) and Drton (2008)], where Σ is bivariate and diagonal whereas the mean vector satisfies the restriction $\mu_1 = \mu_2$.

Kruskal (1968) found the following necessary and sufficient condition for the two estimators to agree for fixed Σ:

THEOREM 1 (Kruskal). *Let Y be a random vector in an inner product space with unknown mean μ in a linear space Ω and known covariance matrix Σ. Then the estimators μ^* and $\hat{\mu}$ coincide if and only if Ω is invariant under $K = \Sigma^{-1}$, that is, if and only if*

$$(4) \qquad\qquad K\Omega \subseteq \Omega.$$

Consequently, if (4) is satisfied by all K in a model, the likelihood function can be maximized for fixed K by μ^*, and inference on K can be based on the profile likelihood (3).

4. Model types.

4.1. *Descriptions.* As stated earlier, we consider three model types introduced in Højsgaard and Lauritzen (2008), which can be represented by colored graphs. These are discussed briefly here, and we refer to Højsgaard and Lauritzen (2008) for further details.

RCON models: Restrictions on concentrations. RCON models place equality constraints on the concentration matrix K. They restrict off-diagonal elements of K separately from those on the diagonal, so that the restrictions can be represented by a graph coloring $(\mathcal{V}, \mathcal{E})$ of G, with $\mathcal{V}$ representing the diagonal and $\mathcal{E}$ the off-diagonal constraints. The corresponding set of positive definite matrices is denoted by $\mathcal{S}^+(\mathcal{V}, \mathcal{E})$.

RCOR models: Restrictions on partial correlations. RCOR models combine equality restrictions on the diagonal of K with equality constraints on partial correlations, given by

$$(5) \qquad \rho_{\alpha\beta|V\setminus\{\alpha,\beta\}} = -\frac{k_{\alpha\beta}}{\sqrt{k_{\alpha\alpha}k_{\beta\beta}}}, \qquad \alpha, \beta \in V, \alpha \neq \beta.$$

Constraints of RCOR models may also be represented by a graph coloring $(\mathcal{V}, \mathcal{E})$, and we denote the corresponding set of positive definite matrices by $\mathcal{R}^+(\mathcal{V}, \mathcal{E})$.

RCOP models: Permutation symmetry. RCOP models are determined by distribution invariance under a group of permutations of the vertices which preserve the edges of the graph, that is, a subgroup of $\mathrm{Aut}(G)$, the group of *automorphisms* of G. For $\sigma \in \mathrm{Aut}(G)$, let $G(\sigma)$ be the permutation matrix representing σ, with $G(\sigma)_{\alpha\beta} = 1$ if and only if σ maps β to α, for $\alpha, \beta \in V$. Then a Gaussian $\mathcal{N}_{|V|}(0, \Sigma)$ distribution is preserved by σ if and only if

$$(6) \qquad G(\sigma)KG(\sigma)^{-1} = K.$$

The RCOP model generated by a group $\Gamma \subseteq \mathrm{Aut}(G)$ assumes

$$K \in \mathcal{S}^+(G, \Gamma) = \mathcal{S}^+(G) \cap \mathcal{S}^+(\Gamma),$$

where $\mathcal{S}^+(\Gamma)$ denotes the set of positive definite matrices satisfying (6) for all $\sigma \in \Gamma$ [Højsgaard and Lauritzen (2008)].

4.2. *Relations between model types.* Under certain conditions on the coloring, RCON and RCOR models, which are determined by the same colored graph, coincide in their model restrictions. First we define edge regularity of a graph coloring.

DEFINITION 1.   Let $\mathcal{G} = (\mathcal{V}, \mathcal{E})$ be a colored graph. We say that $(\mathcal{V}, \mathcal{E})$ is *edge regular* if any pair of edges in the same color class in $\mathcal{E}$ connects the same vertex color classes.

The relevant result in Højsgaard and Lauritzen (2008) then becomes:

PROPOSITION 1.   *The RCON and RCOR models, that are determined by the colored graph $\mathcal{G} = (\mathcal{V}, \mathcal{E})$, yield identical restrictions, that is,*

$$\mathcal{S}^+(\mathcal{V}, \mathcal{E}) = \mathcal{R}^+(\mathcal{V}, \mathcal{E})$$

if and only if $(\mathcal{V}, \mathcal{E})$ is edge regular.

RCOP models fall into the class of models which satisfy the condition in Proposition 1, as we show below:

PROPOSITION 2.   *If a colored graph $\mathcal{G} = (\mathcal{V}, \mathcal{E})$ represents an RCOP model, then $(\mathcal{V}, \mathcal{E})$ is edge regular.*

PROOF.   Let $\mathcal{G} = (\mathcal{V}, \mathcal{E})$ represent an RCOP model, generated by permutation group $\Gamma \subseteq \mathrm{Aut}(G)$, say, and let $u \in \mathcal{E}$ and $e, f \in u$. By definition, there exists $\sigma \in \Gamma$ mapping e to f while leaving $(\mathcal{V}, \mathcal{E})$ invariant. This implies that the colorings of the end vertices of e and f must be identical. $\square$

Thus, if for a graph $G = (V, E)$ we let $\mathcal{V}$ and $\mathcal{E}$ denote the vertex orbits and edge orbits of a group $\Gamma \subseteq \mathrm{Aut}(G)$, then $\mathcal{S}^+(G, \Gamma) = \mathcal{S}^+(\mathcal{V}, \mathcal{E}) = \mathcal{R}^+(\mathcal{V}, \mathcal{E})$ [Højsgaard and Lauritzen (2008)].

For example, since the coloring of the graph in Figure 1 is generated by the group $\Gamma = \{I, \sigma\}$ with σ simultaneously permuting B_1 with B_2 and L_1 with L_2, the corresponding two sets $\mathcal{S}^+(\mathcal{V}, \mathcal{E})$ and $\mathcal{R}^+(\mathcal{V}, \mathcal{E})$ coincide.

5. Equality of maximum likelihood and least-squares estimator.   By Theorem 1, the maximum likelihood estimator $\hat{\mu}$ and least-squares estimator μ^* agree for μ in a linear subspace $\Omega \subseteq \mathbb{R}^V$ if and only if Ω is stable under all K in the model. Below we show that for RCON and RCOR models, and thus also for RCOP models, invariance of Ω under K is equivalent to invariance under $\{T^u\}_{u \in \mathcal{V} \cup \mathcal{E}}$.

PROPOSITION 3.   *Let $\mathcal{G} = (\mathcal{V}, \mathcal{E})$ be a colored graph representing the RCON model with $K \in \mathcal{S}^+(\mathcal{V}, \mathcal{E})$. Then for $\Omega \subseteq \mathbb{R}^V$,*

$$K\Omega \subseteq \Omega \qquad \forall K \in \mathcal{S}^+(\mathcal{V}, \mathcal{E}) \quad \Longleftrightarrow \quad T^u\Omega \subseteq \Omega \qquad \forall u \in \mathcal{V} \cup \mathcal{E}.$$

PROOF.   By definition of RCON models, all $K \in \mathcal{S}^+(\mathcal{V}, \mathcal{E})$ can be written as

$$(7) \qquad\qquad K = \sum_{u \in \mathcal{V} \cup \mathcal{E}} \theta_u T^u, \qquad \theta_u \in \mathbb{R} \text{ for } u \in \mathcal{V} \cup \mathcal{E}$$

with $\{\theta_u\}_{u \in \mathcal{V} \cup \mathcal{E}}$ such that the expression in (7) is positive definite. Suppose first that we have

$$T^u \Omega \subseteq \Omega \qquad \forall u \in \mathcal{V} \cup \mathcal{E}.$$

Since Ω is a linear space this implies invariance under all $K \in \mathcal{S}^+(\mathcal{V}, \mathcal{E})$.

Next suppose $K\Omega \subseteq \Omega$ for all $K \in \mathcal{S}^+(\mathcal{V}, \mathcal{E})$. $\mathcal{S}^+(\mathcal{V}, \mathcal{E})$ is an open convex cone, so that for all $K \in \mathcal{S}^+(\mathcal{V}, \mathcal{E})$ and $u \in \mathcal{V} \cup \mathcal{E}$ there exists $\lambda_u \in \mathbb{R} \setminus \{0\}$ such that

$$K_u = K + \lambda_u T^u \in \mathcal{S}^+(\mathcal{V}, \mathcal{E}).$$

By assumption, $K\Omega \subseteq \Omega$ and $K_u \Omega \subseteq \Omega$, which gives

$$(K_u - K)\Omega = \lambda_u T^u \Omega \subseteq \Omega$$

and thus the desired result. $\square$

Although the cone $\mathcal{R}^+(\mathcal{V}, \mathcal{E})$ is not in general convex, the same holds for RCOR models, as shown below.

PROPOSITION 4. *Let $\mathcal{G} = (\mathcal{V}, \mathcal{E})$ be a colored graph. Then for $\Omega \subseteq \mathbb{R}^V$,*

$$K\Omega \subseteq \Omega \qquad \forall K \in \mathcal{R}^+(\mathcal{V}, \mathcal{E}) \quad \Longleftrightarrow \quad T^u \Omega \subseteq \Omega \qquad \forall u \in \mathcal{V} \cup \mathcal{E}.$$

PROOF. Let $A = (a_{\alpha\beta})_{\alpha,\beta \in V}$ denote the diagonal matrix with entries equal to the inverse partial standard deviations, that is,

$$a_{\alpha\alpha} = \sqrt{k_{\alpha\alpha}} \qquad \text{for } \alpha \in V,$$

and let $C = (c_{\alpha\beta})_{\alpha,\beta \in V}$ have all diagonal entries equal to 1 and all off-diagonal entries be given by the negative partial correlations $-\rho_{\alpha\beta|V \setminus \{\alpha,\beta\}}$. Then, by (5), all $K \in \mathcal{R}^+(\mathcal{V}, \mathcal{E})$ can be uniquely expressed as

$$(8) \qquad K = ACA = \sum_{v,w \in \mathcal{V}} a_v a_w T^v T^w + \sum_{v,w \in \mathcal{V}, u \in \mathcal{E}} c_u a_v a_w T^v T^u T^w,$$

where for $v \in \mathcal{V}$ and $u \in \mathcal{E}$, we let a_v and c_u denote $a_{\alpha\alpha}$, $\alpha \in v$ and $c_{\alpha\beta}$, $\alpha\beta \in u$, respectively. As for $v, w \in \mathcal{V}$, $T^v T^w$ is zero unless $v = w$, when it equals T^v, and equation (8) simplifies to

$$(9) \qquad K = \sum_{v \in \mathcal{V}} a_v^2 T^v + \sum_{v,w \in \mathcal{V}, u \in \mathcal{E}} c_u a_v a_w T^v T^u T^w.$$

Suppose first that we have

$$T^u \Omega \subseteq \Omega \qquad \forall u \in \mathcal{V} \cup \mathcal{E}.$$

As before, since Ω is a linear space, (9) implies invariance under all $K \in \mathcal{R}^+(\mathcal{V}, \mathcal{E})$. So suppose $K\Omega \subseteq \Omega$ for all $K \in \mathcal{R}^+(\mathcal{V}, \mathcal{E})$. Then, in particular, Ω

is invariant under all K of the form

$$K = ACA \qquad \text{with } A = \sigma^2 I \text{ and } C \text{ as specified above,}$$

represented by the graph coloring $(\{V\}, \mathcal{E})$, which has the same edge coloring as $(\mathcal{V}, \mathcal{E})$, but with all vertices of the same color. This graph coloring is clearly edge regular (as all edges connect the only vertex color class back to itself), which gives that the represented model is also of type RCON. We may therefore apply Proposition 3, giving

$$T^u \Omega \subseteq \Omega \qquad \forall u \in \mathcal{E}.$$

For the vertex coloring, consider the submodel

$$K = ACA \qquad \text{with } A \text{ as specified above and } C = I.$$

This submodel is represented by the independence graph with no edges and vertex coloring $\mathcal{V}$, which is also edge regular, so that by Proposition 3,

$$T^v \Omega \subseteq \Omega \qquad \forall v \in \mathcal{V},$$

completing the proof. $\square$

5.1. *Equality restrictions in the means of RCON and RCOR models.* Let $\mathcal{M}$ be a partition of V and consider $\Omega(\mathcal{M})$ as in (1). In the following we derive a necessary and sufficient condition on $\mathcal{M}$ and $(\mathcal{V}, \mathcal{E})$ for $\Omega(\mathcal{M})$ to be invariant under all $K \in \mathcal{S}^+(\mathcal{V}, \mathcal{E})$ or all $K \in \mathcal{R}^+(\mathcal{V}, \mathcal{E})$. By Propositions 3 and 4 we may consider vertex and edge color classes of the colored graph representing the model separately. We begin with the vertex coloring.

PROPOSITION 5.   *Let $\mathcal{M}$ and $\mathcal{V}$ be partitions of V. Then $\Omega(\mathcal{M})$ is invariant under $\{T^v\}_{v \in \mathcal{V}}$ if and only if $\mathcal{M} \leq \mathcal{V}$.*

PROOF.   The action of T^v for $v \in \mathcal{V}$ on $\mu \in \mathbb{R}^V$ is given as

$$T^v \mu = \begin{cases} \mu_\alpha & \text{if } \alpha \in v, \\ 0 & \text{otherwise.} \end{cases}$$

Let $\mu \in \Omega(\mathcal{M})$, and suppose that $T^v \Omega(\mathcal{M}) \subseteq \Omega(\mathcal{M})$ for all $v \in \mathcal{V}$. In order for $T^v \mu \in \Omega(\mathcal{M})$ for all $v \in \mathcal{V}$, we must have $\alpha \equiv \beta(\mathcal{V})$ whenever $\alpha \equiv \beta(\mathcal{M})$, or equivalently $\mathcal{M} \leq \mathcal{V}$. Conversely, suppose $\mathcal{M} \leq \mathcal{V}$. Then $\alpha \equiv \beta(\mathcal{V})$ whenever $\alpha \equiv \beta(\mathcal{M})$, which gives $T^v \mu \in \Omega(\mathcal{M})$ for all $v \in \mathcal{V}$. $\square$

Note that the above result implies that the likelihood cannot be maximized in μ independently of the value of K in the Behrens–Fisher setting, which is the RCON and RCOR model on two variables specified by $\mathcal{V} = \{\{1\}, \{2\}\}$, $\mathcal{E} = \varnothing$ together with the restriction $\mu_1 = \mu_2$ on the means, as the mean partitioning is then coarser than the vertex coloring.

Also, for the model of Frets's heads in Figure 1, the MLE of the mean is not simple if we, for example, wish to estimate the mean under the hypothesis that the heads tend to be square-shaped, that is, if mean lengths are equal to mean breadths, as this partition would not be finer than the vertex coloring.

For the edge coloring, we require the concept of an *equitable partition*, first defined by Sachs (1966).

DEFINITION 2 (Sachs). Let $G = (V, E)$ be an undirected graph. Then a partition, or equivalent coloring, $\mathcal{V}$ of V is called *equitable with respect to G* if for all $v \in \mathcal{V}$, $\alpha, \beta \in v$,

$$|\operatorname{ne}_E(\alpha) \cap w| = |\operatorname{ne}_E(\beta) \cap w| \qquad \forall w \in \mathcal{V}.$$

Chan and Godsil (1997) proved the following.

PROPOSITION 6 (Chan and Godsil). *Let $G = (V, E)$ be an undirected graph with adjacency matrix T, and let $\mathcal{M}$ be a partition of V. Then $\Omega(\mathcal{M})$ is invariant under T if and only if $\mathcal{M}$ is equitable with respect to G.*

The notion of an equitable partition for vertex colored graphs can be naturally extended to graphs with colored vertices and edges. We term the corresponding graph colorings *vertex regular*, defined below.

DEFINITION 3. Let $\mathcal{G} = (\mathcal{V}, \mathcal{E})$ be a colored graph, and let the subgraph induced by the edge color class $u \in \mathcal{E}$ be denoted by $G^u = (V, u)$. We say that $(\mathcal{V}, \mathcal{E})$ is *vertex regular* if $\mathcal{V}$ is equitable with respect to G^u for all $u \in \mathcal{E}$.

Combining Definition 3 with Proposition 6 yields vertex regularity to be a necessary and sufficient condition for $\Omega(\mathcal{M})$ to be invariant under $\{T^u\}_{u \in \mathcal{E}}$:

PROPOSITION 7. *For $\mathcal{G} = (\mathcal{V}, \mathcal{E})$ a colored graph and $\mathcal{M}$ a partition of V, $\Omega(\mathcal{M})$ is invariant under $\{T^u\}_{u \in \mathcal{E}}$ if and only if $(\mathcal{M}, \mathcal{E})$ is vertex regular.*

PROOF. For $u \in \mathcal{E}$, T^u is the adjacency matrix of G^u. By Proposition 6, $\Omega(\mathcal{M})$ is stable under $\{T^u\}_{u \in \mathcal{E}}$ if and only if $\mathcal{M}$ is equitable with respect to G^u for all $u \in \mathcal{E}$, or equivalently if and only if $(\mathcal{M}, \mathcal{E})$ is vertex regular. □

For the Frets's heads model in Figure 1, this implies that restricting the mean breadths and mean lengths on their own does not ensure $\hat{\mu} = \mu^*$, as the corresponding partitions $\mathcal{M} = \{\{B_1, B_2\}, \{L_1\}, \{L_2\}\}$ and $\mathcal{M} = \{\{B_1\}, \{B_2\}, \{L_1, L_2\}\}$ do not give rise to vertex regular colorings $(\mathcal{M}, \mathcal{E})$.

Combining Proposition 5 and Proposition 7 establishes our main result:

THEOREM 2. *Let $\mathcal{G} = (\mathcal{V}, \mathcal{E})$ be a colored graph, let $\mathcal{M}$ be a partition of V and consider a sample from a multivariate Gaussian $\mathcal{N}_{|V|}(\mu, \Sigma)$ distribution*

with $\mu \in \Omega(\mathcal{M})$ and $K = \Sigma^{-1} \in \mathcal{S}^+(\mathcal{V}, \mathcal{E})$, both unknown. It then holds that

(10) $\hat{\mu} = \mu^* \iff \mathcal{M} \leq \mathcal{V}$ and $(\mathcal{M}, \mathcal{E})$ is vertex regular.

The same conclusion holds if $\mathcal{S}^+(\mathcal{V}, \mathcal{E})$ is replaced by $\mathcal{R}^+(\mathcal{V}, \mathcal{E})$.

In fact, for any colored graph $\mathcal{G} = (\mathcal{V}, \mathcal{E})$ there is always a coarsest partition $\mathcal{M}$ satisfying the conditions in (10) [Gehrmann (2011)]. The finest variant of such a coarsest equitable refinement $\mathcal{M}$ of $\mathcal{V}$ is given by the singleton partition. Clearly it is finer than any vertex coloring, and further naturally gives a vertex regular coloring $(\mathcal{M}, \mathcal{E})$. Note that in this case $\Omega(\mathcal{M})$ corresponds to the unrestricted case considered in Section 3, conforming with the fact that then $\hat{\mu} = \mu^*$.

5.2. *Equality restrictions in the means in RCOP models.*   The coarsest possible $\mathcal{M}$ for RCON and RCOR models, by (10), is $\mathcal{V}$, for which $\hat{\mu} = \mu^*$ if and only if $(\mathcal{V}, \mathcal{E})$ is vertex regular. This always holds for RCOP models.

PROPOSITION 8.   *If a colored graph $\mathcal{G} = (\mathcal{V}, \mathcal{E})$ represents an RCOP model, then $(\mathcal{V}, \mathcal{E})$ is vertex regular.*

PROOF.   Let $\mathcal{G} = (\mathcal{V}, \mathcal{E})$ represent an RCOP model, generated by a subgroup $\Gamma \subseteq \mathrm{Aut}(G)$, say. By Proposition 2, $(\mathcal{V}, \mathcal{E})$ is edge regular. Thus whenever two edges $e, f \in E$ are of the same color, they connect the same vertex color classes. Let $\alpha, \beta \in V$ be two equally colored vertices in $\mathcal{V}$. Then, by definition of RCOP models, there exists a permutation $\sigma \in \Gamma$ which maps α to β leaving $(\mathcal{V}, \mathcal{E})$ invariant. This implies that the degree in each edge color class of α and β must be identical. The previous two statements imply

$$|\mathrm{ne}_u(\alpha) \cap v| = |\mathrm{ne}_u(\beta) \cap v|$$

for all $v \in \mathcal{V}$ and all pairs $\alpha, \beta \in V$ with $\alpha \equiv \beta(\mathcal{V})$, which is precisely the criterion of vertex regularity for the graph coloring $(\mathcal{V}, \mathcal{E})$. $\square$

We mention in passing that Proposition 2 and Proposition 8 combined establish that colorings of graphs which represent RCOP models are *regular* in the terminology of Siemons (1983). We conclude from Theorem 2 and Proposition 8:

COROLLARY 1.   *Let $G = (V, E)$ be an undirected graph, and let $(\mathcal{V}, \mathcal{E})$ represent the constraints of an RCOP model generated by group $\Gamma \subseteq \mathrm{Aut}(G)$. Then for a sample from a multivariate Gaussian $\mathcal{N}_{|V|}(\mu, \Sigma)$ distribution with $\mu \in \Omega(\mathcal{V})$ and $K = \Sigma^{-1} \in \mathcal{S}^+(G, \Gamma)$, both unknown, we always have $\hat{\mu} = \mu^*$.*

6. Examples.   We first consider the example in Figure 1 on head dimensions for first and second sons. Representing an RCOP model, by Proposition 8, the graph in Figure 1 has a vertex regular coloring. It follows from Corollary 1 that the maximum likelihood estimate of the mean under the
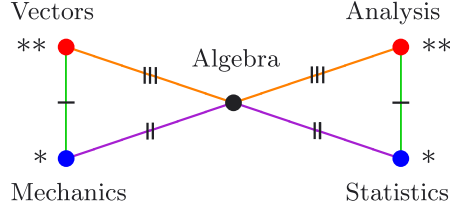
Fig. 2. *Mathematics marks example.*

hypothesis that the mean length and mean breadth are equal for the two sons is simply the total average of the head lengths and the head breadths, respectively. Similarly, it follows from Theorem 2 that the only hypotheses about the mean that have a simple solution are this one and the one where the means are completely unrestricted.

The empirical means of the dimensions (B_1, B_2, L_1, L_2) are equal to $(151.12, 149.24, 185.72, 183.84)$, so that the MLE of the means under the hypothesis that the mean lengths and breadths are independent of the parity of the son then become $(150.18, 150.18, 184.78, 184.78)$. The likelihood ratio test is obtained by comparing the maximized profile likelihoods (3) calculated with appropriate residual covariance matrices W under the two hypotheses. Using the R-package gRc [Højsgaard and Lauritzen (2011)] this yields $-2 \log \mathrm{LR} = 3.27$ on 2 degrees of freedom, so there is no evidence for the sizes depending on the parity of the son.

Our second example is concerned with the examination marks of 88 students in five mathematical subjects [Mardia, Kent and Bibby (1979)]. The RCOP model represented by $\mathcal{G} = (\mathcal{V}, \mathcal{E})$ in Figure 2 was demonstrated to be an excellent fit in Højsgaard and Lauritzen (2008).

The model is given by invariance of K under simultaneously permuting Mechanics with Statistics, and Vectors with Analysis. In their fit, Højsgaard and Lauritzen (2008) implicitly assumed an unconstrained mean. The MLE $\hat{\mu}$ is then given by the sample averages $(\hat{\mu}_{\mathrm{al}}, \hat{\mu}_{\mathrm{an}}, \hat{\mu}_{\mathrm{me}}, \hat{\mu}_{\mathrm{st}}, \hat{\mu}_{\mathrm{ve}}) = (50.60, 46.68, 38.96, 42.31, 50.59)$ in the obvious notation, which corresponds to $\mathcal{M}$ being the singleton partition. However, it could be natural to assume μ subject to the same invariance as K, meaning $\mathcal{M} = \mathcal{V}$, or in this case $\mu_{\mathrm{an}} = \mu_{\mathrm{ve}}$ and $\mu_{\mathrm{me}} = \mu_{\mathrm{st}}$. Then, by Corollary 1, $(\hat{\mu}_{\mathrm{al}}, \hat{\mu}_{\mathrm{an}}, \hat{\mu}_{\mathrm{me}}, \hat{\mu}_{\mathrm{st}}, \hat{\mu}_{\mathrm{ve}}) = (50.60, 48.64, 40.63, 40.63, 48.64)$. The likelihood ratio statistic for this mean structure relative to the model in Figure 2 with unconstrained mean takes the value of 11.9 on 2 degrees of freedom, and the hypothesis about symmetry in the means is therefore clearly rejected with $p < 0.003$.

Note that the sample averages for Vectors and Algebra are almost identical. However, combining the constraints on K represented by the graph in Figure 2 with any hypothesis on the means implying $\mu_{\mathrm{ve}} = \mu_{\mathrm{al}}$ will require joint maximization in μ and K of the likelihood function in (2) to obtain $\hat{\mu}$.

7. Discussion. The main result of this article is a necessary and sufficient condition on the pattern of equality constraints on the mean vector μ in a graphical Gaussian symmetry model with colored graph $\mathcal{G} = (\mathcal{V}, \mathcal{E})$ which ensures the identity of the least squares estimate μ^* and the maximum likelihood estimate $\hat{\mu}$, given in Theorem 2.

The derived necessary and sufficient condition is formulated in terms of vertex and edge colored graphs and is easily testable, so that any set of equality constraints on μ together with constraints on either K or the partial correlations can be verified for estimability of μ by μ^*.

The result is, for example, useful if one set of constraints, either on the mean or the independence structure, is assumed to be given and the other may be varied. A setting which falls into this category is the design of experiments seeking to estimate mean treatment effects under the assumption of correlations with inherent symmetries in the error structure at experimental sites. A systematic arrangement of the sites may enforce a symmetry pattern in the concentrations or partial correlations, and thus restrict the concentration matrix as in one of models considered here. Allocation of treatments then effectively restricts the mean response at sites with the same treatment to be identical, and the condition derived can be used to find treatment allocations which ensure estimability of mean treatment effects without knowledge of the value of Σ.

An interesting question directly emerging from our work is concerned with the exact distributions of likelihood ratio test statistics for hypotheses about the mean in RCON, RCOR and RCOP models. In the examples discussed in this paper we have relied on asymptotic theory to judge the significance of test statistics, but it is likely that their distributions can be derived explicitly, for example when the mean hypothesis is given by the natural symmetry of an RCOP model. Hylleberg, Jensen and Ørnbøl (1993) developed explicit likelihood ratio tests for decomposable mean-zero RCOP models generated by compound symmetry [Votaw (1948)], and it would be interesting to extend these results to models with nonzero means and more general symmetry constraints.

We note that if our condition is not satisfied, we would expect phenomena similar to those in the Behrens–Fisher problem implying the general nonuniqueness of the MLE [Drton (2008)] and the nonexistence of an α-similar test of the mean hypothesis [Scheffé (1944)].

REFERENCES

ANDERSEN, H. H., HØJBJERRE, M., SØRENSEN, D. and ERIKSEN, P. S. (1995). *Linear and Graphical Models for the Multivariate Complex Normal Distribution. Lecture Notes in Statistics* **101**. Springer, New York. MR1440852

ANDERSSON, S. (1975). Invariant normal models. *Ann. Statist.* **3** 132–154. MR0362703

ANDERSSON, S. A., BRØNS, H. K. and JENSEN, S. T. (1983). Distribution of eigenvalues in multivariate statistical analysis. *Ann. Statist.* **11** 392–415. MR0696055

BOLLOBÁS, B. (1998). *Modern Graph Theory. Graduate Texts in Mathematics* **184**. Springer, New York. MR1633290

CHAN, A. and GODSIL, C. D. (1997). Symmetry and eigenvectors. In *Graph Symmetry (Montreal, PQ, 1996). NATO Adv. Sci. Inst. Ser. C Math. Phys. Sci.* **497** 75–106. Kluwer Academic, Dordrecht. MR1468788

DAWID, A. P. (1988). Symmetry models and hypotheses for structured data layouts. *J. Roy. Statist. Soc. Ser. B* **50** 1–34. MR0954729

DIACONIS, P. (1988). *Group Representations in Probability and Statistics. Institute of Mathematical Statistics Lecture Notes—Monograph Series* **11**. IMS, Hayward, CA. MR0964069

DRTON, M. (2008). Multiple solutions to the likelihood equations in the Behrens–Fisher problem. *Statist. Probab. Lett.* **78** 3288–3293. MR2479491

EATON, M. L. (1983). *Multivariate Statistics: A Vector Space Approach.* Wiley, New York. MR0716321

EATON, M. L. (1989). *Group Invariance Applications in Statistics. Regional Conference Series in Probability and Statistics* **1**. IMS and American Statistical Association, Hayward, CA and Alexandria, VA. MR1089423

FRETS, G. P. (1921). Heredity of head form in man. *Genetica* **3** 193–400.

GEHRMANN, H. (2011). Lattices of graphical Gaussian models with symmetries. *Symmetry* **3** 653–679. MR2845303

HABERMAN, S. J. (1975). How much do Gauss–Markov and least square estimates differ? A coordinate-free approach. *Ann. Statist.* **3** 982–990. MR0378269

HØJSGAARD, S. and LAURITZEN, S. L. (2008). Graphical Gaussian models with edge and vertex symmetries. *J. R. Stat. Soc. Ser. B Stat. Methodol.* **70** 1005–1027. MR2530327

HØJSGAARD, S. and LAURITZEN, S. L. (2011). gRc: Inference in graphical Gaussian models with edge and vertex symmetries. R package version 0.3.0.

HYLLEBERG, B., JENSEN, M. and ØRNBØL, E. (1993). Graphical symmetry models. M.Sc. thesis, Aalborg Univ., Aalborg.

JENSEN, S. T. (1988). Covariance hypotheses which are linear in both the covariance and the inverse covariance. *Ann. Statist.* **16** 302–322. MR0924873

KRUSKAL, W. (1968). When are Gauss–Markov and least squares estimators identical? A coordinate-free approach. *Ann. Math. Statist* **39** 70–75. MR0222998

LAURITZEN, S. L. (1996). *Graphical Models. Oxford Statistical Science Series* **17**. Clarendon Press, New York. MR1419991

MADSEN, J. (2000). Invariant normal models with recursive graphical Markov structure. *Ann. Statist.* **28** 1150–1178. MR1810923

MARDIA, K. V., KENT, J. T. and BIBBY, J. M. (1979). *Multivariate Analysis.* Academic Press, New York, NY.

OLKIN, I. (1972). Testing and estimation for structures which are circularly symmetric in blocks. Technical report, Educational Testing Service, Princeton, NJ.

OLKIN, I. and PRESS, S. J. (1969). Testing and estimation for a circular stationary model. *Ann. Math. Statist.* **40** 1358–1373. MR0245139

SACHS, H. (1966). Über Teiler, Faktoren und charakteristische Polynome von Graphen. *Wissenschaftliche Zeitschrift der Technischen Hochschule Ilmenau* **12** 7–12.

SCHEFFÉ, H. (1944). A note on the Behrens–Fisher problem. *Ann. Math. Statist.* **15** 430–432. MR0011926

SIEMONS, J. (1983). Automorphism groups of graphs. *Arch. Math. (Basel)* **41** 379–384. MR0731610

VIANA, M. A. G. (2008). *Symmetry Studies: An Introduction to the Analysis of Structured Data in Applications.* Cambridge Univ. Press, Cambridge. MR2419845

VOTAW, D. F. JR. (1948). Testing compound symmetry in a normal multivariate distribution. *Ann. Math. Statist.* **19** 447–473. MR0027999

WHITTAKER, J. (1990). *Graphical Models in Applied Multivariate Statistics.* Wiley, Chichester. MR1112133

WILKS, S. S. (1946). Sample criteria for testing equality of means, equality of variances, and equality of covariances in a normal multivariate distribution. *Ann. Math. Statist.* **17** 257–281. MR0017498

DEPARTMENT OF STATISTICS
UNIVERSITY OF OXFORD
1 SOUTH PARKS ROAD
OXFORD OX1 3TG
UNITED KINGDOM
E-MAIL: gehrmann@stats.ox.ac.uk
          steffen@stats.ox.ac.uk